Research Article

Whitening of Background Brain Activity via Parametric Modeling

Nidal Kamel, Andrews Samraj, and Arash Mousavi

Several signal subspace techniques have been recently suggested for the extraction of the visual evoked potential signals from brain background colored noise. The majority of these techniques assume the background noise as white, and for colored noise, it is suggested to be whitened, without further elaboration on how this might be done. In this paper, we investigate the whitening capabilities of two parametric techniques: a direct one based on Levinson solution of *Yule-Walker* equations, called AR *Yule-Walker*, and an indirect one based on the least-squares solution of *forward-backward linear prediction* (FBLP) equations, called AR-FBLP. The whitening effect of the two algorithms is investigated with real background electroencephalogram (EEG) colored noise and compared in time and frequency domains.

1. Introduction

The detection of steady-state visual evoked potentials (VEPs) is important in some clinical audiometry and ophthalmology applications [1]. The VEPs are usually concealed in the ongoing background electroencephalogram (EEG) generated in the brain. The EEG is highly colored with unknown covariance matrix. Several subspace-based techniques were proposed to extract the VEPs signal from the background colored noise. The majority of these techniques assume the background noise as white, and for colored noise, it is suggested to be whitened, without further elaboration on how this might be done [2]. In this paper, we investigate the capabilities of the autoregressive (AR) model of random process in whitening the EEG colored noise. The AR-based whitening process consists of two

major steps: (1) estimate the autoregressive model parameters that best fit the EEG data signal, (2) use the EEG data as input to *error linear predictor* (ELP) filter of similar order and coefficients to AR. Between the aforementioned two steps in whitening the EEG colored noise, step one remains of greater importance. During the last three decades, several methods have been proposed to estimate the autoregressive model parameters from the available data sequence [3, 4]. Though the great diversity of the proposed methods, the majority of them fall into the following two groups: (1) direct AR parameter estimates, which are based on Levinson recursive solution to Yule-Walker equations, (2) indirect AR parameter estimates, which first make estimates of the linear prediction coefficients and equate them later with AR parameters. Generally speaking, the direct methods presume that the second moment statistics of the random process are known and develop the equations whose solution provides the model parameters; but in most practical applications, the fixed quantities appearing in Yule-Walker equations, namely the values of the correlation function, are not known a priori but need to be estimated from data. This introduces errors and various other complications and brings up the whole issue of estimation of the model parameters [5, 6]. In the contrary to the direct methods in estimating AR parameters, the indirect are based directly on the observed data and seek to minimize a sum of squares of the errors of linear prediction, thus the procedures are referred to as least-squares methods.

In this paper, we investigate the whitening capabilities of two parametric techniques: a direct one based on Levinson solution of *Yule-Walker* equations, called AR *Yule-Walker*, and an indirect one based on the least-squares solution of *forward-backward linear prediction* (FBLP) equations, called AR-FBLP.

To make the presentation as clear as possible, an attempt is made to adhere to somewhat standard notational convention. Lowercase boldface character will generally refer to vector. Uppercase boldface character will generally refer to matrices. For either real or complex-valued matrices, $[\cdot]^H$ will be used to denote the Hermitian conjugate, and $[\cdot]^T$ will denote the matrix transpose operation.

2. The AR Yule-Walker

As stated earlier, the VEPs are embedded in background colored EEG noise of unknown covariance matrix. Though the EEG is a nonstationary random process, it could be assumed stationary over short time intervals (Sections 4 and 5). In order to whiten the colored EEG noise, one must model it as an autoregressive process of certain order (p) and unknown parameters $\{a(1), a(2),\ldots,a(p)\}$:

$$x(n) = -\sum_{k=1}^{p} a(k)x(n-k) + u(n), \tag{2.1}$$

where $u(n)$ is a zero-mean white Gaussian noise of variance σ^2, called driving noise, and $x(n)$ is the EEG signal. If (2.1) is multiplied by $x^*(n-m)$ and the expectation is taken,

then the result is

$$r_{xx}(m) = \begin{cases} -\sum_{k=1}^{p} a(k)r_{xx}(m-k) & \text{for } m > 0, \\ -\sum_{k=1}^{p} a(k)r_{xx}(-k) + \rho_w & \text{for } m = 0, \\ r_{xx}^*(-m) & \text{for } m < 0, \end{cases} \tag{2.2}$$

where $r_{xx}(m)$ is the autocorrelation sample for lag m. This relationship may be evaluated for $p+1$ lag indices $0 \le m \le p$ and formed into the matrix expression

$$\begin{bmatrix} r_{xx}(0) & r_{xx}(-1) & \cdots & r_{xx}(-p) \\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(-p+1) \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}(p) & r_{xx}(p-1) & \cdots & r_{xx}(0) \end{bmatrix} \begin{bmatrix} 1 \\ a(1) \\ \vdots \\ a(p) \end{bmatrix} = \begin{bmatrix} \rho_w \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{2.3}$$

If given, therefore, the autocorrelation sequence for lags 0 to p, then the AR parameters may be found as the solution to (2.3). This relationship forms the AR Yule-Walker equations. The autocorrelation matrix of (2.3) is both Toeplitz and Hermitian, thus *Levinson* algorithm can be used to solve it [7, 8].

In order to understand how an autoregressive model of the EEG colored noise can be used to whiten it, we must understand first the relationship between AR model and linear predictor (LP).

2.1. Relationship of AR to linear prediction analysis. Consider the forward linear prediction estimate

$$\hat{x}^f(n) = -\sum_{k=1}^{m} a^f(k)x(n-k) \tag{2.4}$$

of the sample $x(n)$, where $a^f(k)$ is the forward linear prediction coefficients at time index k, and m is the predictor order. The hat ˆ is used to denote an estimate and the superscript f is used to denote that this is a forward estimate. The prediction is forward in the sense that the estimate at time index n is based on m samples indexed earlier in time. The complex forward linear prediction error is given as

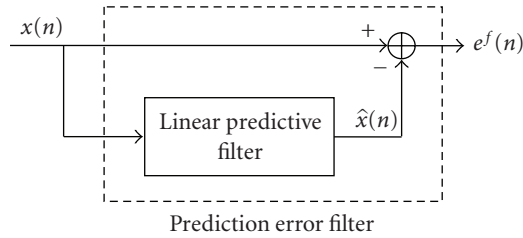$$e^f(n) = x(n) - \hat{x}^f(n) = x(n) + \sum_{k=1}^{m} a^f(k)x(n-k). \tag{2.5}$$

FIGURE 2.1. The implementation of the linear predictor as a whitening filter of an input AR process.

If one rearranges (2.5) as

$$x(n) = -\sum_{k=1}^{m} a^f(k)x(n-k) + e^f(n), \tag{2.6}$$

then the similarity with (2.1), which describes an autoregressive process, may be seen. There are, however, two distinctions between the forward linear prediction process and the AR process. The sequence $u(n)$ of (2.1) is a white noise process used as the input to an autoregressive filter. The sequence $x(n)$ is the output of the autoregressive filter. The error sequence $e^f(n)$ of (2.5) is the output of a forward linear prediction error filter. The sequence $x(n)$ is the input to the prediction error filter. The linear prediction error sequence will be uncorrelated with the linear prediction estimate $\hat{x}^f(n)$, but it is not, in general, a white process unless $x(n)$ was generated as an AR(p) process and $m = p$. The error sequence in this case will be a white process, the forward linear prediction coefficients will be identical to the AR parameters, and the prediction error filter may be regarded as a whitening filter [4].

Figure 2.1 shows the linear predictor as whitening filter of the input AR process.

Based on above discussion, and after solving Yule-Walker equations for the AR(p) coefficients, the EEG colored noise is whitened by passing it through an *error linear predictor* filter (ELP) of similar order and coefficients to AR(p).

The transfer function of the ELP filter is given by

$$A(z) = 1 + a(1)z^{-1} + a(2)z^{-2} + \cdots + a(p)z^{-p}. \tag{2.7}$$

As stated earlier, the second-order quantities in Yule-Walker equations are not known a priori but need to be estimated from data. This introduces errors and various other complications. The alternative is to develop data-oriented technique for AR parameters estimation. Between the data-oriented AR parameters estimates, the forward-backward linear prediction remains among the best.

3. The AR forward-backward linear prediction (FBLP)

Assume that the N-point data sequence $x(1), x(2), \ldots, x(N)$ is to be used to estimate the pth-order forward LP parameters. The forward linear prediction error may be defined

over the range from $n = 1$ to $n = N + p$, if one assumes that the data prior to first sample and after the last sample are zero; but in order to avoid the unrealistic zero assumption about the unavailable data, let us confine the range of prediction errors to $e_p^f(p + 1), e_p^f(p+2), \ldots, e_p^f(N)$. This case is called the *nonwindowed* case, because only available data samples are used, and termed the *covariance* method of linear prediction [9]. The $N - p - 1$ forward linear prediction error terms represented by (2.5) may be summarized, using matrix-vector notation, as

$$
\begin{bmatrix} e^f(p+1) \\ \vdots \\ e^f(N-p) \\ \vdots \\ e^f(N) \end{bmatrix} = \begin{bmatrix} x(p+1) & \cdots & x(1) \\ \vdots & \ddots & \vdots \\ x(N-p) & \cdots & x(p+1) \\ \vdots & \ddots & \vdots \\ x(N) & \cdots & x(N-p) \end{bmatrix} \begin{bmatrix} 1 \\ a^f(1) \\ \vdots \\ a^f(p) \end{bmatrix}.
\tag{3.1}
$$

The $N - p - 1$ forward linear prediction errors in (3.1) may be summarized by the matrix-vector product

$$
\mathbf{e}^f = \mathbf{X}^f \mathbf{a}^f,
\tag{3.2}
$$

where $\mathbf{e}^f$ represents the forward linear prediction error vector, $\mathbf{X}^f$ the forward data matrix, and $\mathbf{a}^f$ the forward linear prediction coefficients.

The forward linear prediction squared error magnitude to be minimized by the coefficients of the filter is simply

$$
\rho^f = \sum_{n=p+1}^{N} |e^f(n)|^2.
\tag{3.3}
$$

Before developing a least-squares solution to the forward filter parameters, it seems reasonable to generate more error points by including the prediction errors in backward direction. The net result should be an improved estimate of the autoregressive parameters [7].

The backward linear predictor for sample $x(n)$ will have the usual form

$$
\hat{x}^b(n) = -\sum_{k=1}^{p} a^b(k)x(n+k),
\tag{3.4}
$$

where $a^b(k)$ are the backward linear prediction coefficients for order p. The prediction is "backward" in the sense that the prediction for the current data sample is a weighted sum

of p future samples. The backward linear prediction error is

$$e^b(n) = x(n) - \hat{x}^b(n) = x(n) + \sum_{k=1}^{m} a^b(k)x(n+k). \tag{3.5}$$

It would be useful to indicate that the backward and forward linear prediction error variances are identical ($\rho^f = \rho^b$) and the backward linear prediction coefficients are simply the complex conjugate of the forward linear prediction coefficients ($a^b(k) = [a^f(k)]^*$) when both prediction filters are the same length p. If the forward linear prediction error filter is a whitening filter of an AR process, then the backward linear prediction error filter is a whitening filter for the anticausal realization of the AR process [8]. Thus, the backward linear prediction error in (3.5) may be rewritten in terms of forward linear predictor coefficients as

$$e^{b*}(n) = x(n)^* + \sum_{k=1}^{m} a^f(k)x(n+k)^*. \tag{3.6}$$

In similar way to the *nonwindowed* forward case, the $N - p - 1$ backward linear prediction error terms represented by (3.6) may be summarized, using matrix-vector notation, as

$$\begin{bmatrix} e^{b*}(1) \\ \vdots \\ e^{b*}(p+1) \\ \vdots \\ e^{b*}(N-p+1) \end{bmatrix} = \begin{bmatrix} x^*(1) & x^*(2) & \cdots & x^*(p+1) \\ \vdots & \ddots & & \vdots \\ x^*(p+1) & x^*(p+2) & \cdots & x^*(N-p) \\ \vdots & \ddots & & \vdots \\ x^*(N-p+1) & x^*(N-p+2) & \cdots & x^*(N) \end{bmatrix} \begin{bmatrix} 1 \\ a^f(1) \\ \vdots \\ a^f(p) \end{bmatrix}. \tag{3.7}$$

The $N - p - 1$ backward linear prediction errors in (3.7) may be summarized by the matrix-vector product

$$\mathbf{e}^b = \mathbf{X}^b \mathbf{a}^f, \tag{3.8}$$

where $\mathbf{e}^b$ represents the forward linear prediction error vector, $\mathbf{X}^b$ the backward data matrix, and $\mathbf{a}^f$ the forward linear prediction coefficients.

The backward linear prediction squared error magnitude to be minimized by the coefficients of the filter is simply

$$\rho^b = \sum_{n=1}^{N-p+1} |e^b(n)|^2. \tag{3.9}$$

As stated earlier, because both directions have similar statistical information, it seems reasonable to combine the linear prediction error statistics of both forward and backward

errors in order to generate more error points. The $(N - p + 1)$ forward and $(N - p + 1)$ backward linear prediction errors of the windowed case can be summarized concisely by the matrix-vector product

$$\mathbf{e} = \begin{bmatrix} \mathbf{e}^f \\ \mathbf{e}^{b*} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^f \\ \mathbf{X}^b \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{a}^{fb} \end{bmatrix}. \tag{3.10}$$

The linear prediction coefficient vector $\mathbf{a}^{fb}$ in (3.10) is given by

$$\mathbf{a}^{fb} = \begin{bmatrix} a(1) \\ \vdots \\ a(p) \end{bmatrix}. \tag{3.11}$$

Observe that the superscript f has been omitted from the linear prediction coefficients $a(k)$ because these apply now to both forward and backward prediction errors.

Minimizing the average of the forward and backward linear prediction squared errors over the available data

$$\rho^{fb} = \frac{1}{2} \left[\sum_{n=p+1}^{N} |e^f(n)|^2 + \sum_{n=1}^{N-p+1} |e^b(n)|^2 \right] = \frac{1}{2} \mathbf{e}^H \mathbf{e} = \frac{1}{2} \left[(\mathbf{e}^f)^H \mathbf{e}^f + (\mathbf{e}^b)^H \mathbf{e}^b \right] \tag{3.12}$$

leads to the set of normal equations

$$\mathbf{R}\,\mathbf{a}^{fb} = \begin{bmatrix} 2\rho^{fb} \\ \mathbf{0}_p \end{bmatrix} \tag{3.13}$$

in which

$$\mathbf{R} = \begin{bmatrix} \mathbf{X}^f \\ \mathbf{X}^b \end{bmatrix}^H \begin{bmatrix} \mathbf{X}^f \\ \mathbf{X}^b \end{bmatrix} \tag{3.14}$$

and $\mathbf{0}_p$ is a p-element all-zeros vector. Solution of these produces the filter coefficients and the combined least-squares error. Matrix $\mathbf{R}$ is not Toeplitz but does have a special structure. Marple has developed a fast way to solve these equations that exploit that structure [7].

After defining the linear predictor coefficients that best model the EEG data segments in prediction sense, they are used as the AR parameters estimates by equating the AR parameters to the forward linear prediction coefficients. The EEG data is then whitened by passing it through an *error linear predictor* (ELP) filter of similar transfer function to (2.7).

4. Model order selection

Because the best choice of filter order p is not generally known a priori, it is usually necessary in practice to estimate it. Many criteria have been proposed as allegedly objective

functions for selection of the AR model order. Akaike [10] has provided two criteria. His first criterion is the *final prediction error* (FPE) given as

$$\text{FPE}(p) = \hat{\rho}_p \left(\frac{N + (p+1)}{N - (p+1)} \right), \tag{4.1}$$

where N is the number of data samples, p is the order, and $\hat{\rho}_p$ is the estimated white noise variance (the linear prediction error variance will be used for this estimate). The order p selected is the one for which the FPE is minimum.

Akaike also suggested another selection criterion using the maximum likelihood approach to derive a criterion termed the Akaike information criterion (AIC) [11]. Assuming that the process has Gaussian statistics, the AIC for AR processes has the form

$$\text{AIC}(p) = N \ln \left(\hat{\rho}_p \right) + 2p. \tag{4.2}$$

Again, the order p selected is the one that minimizes the AIC.

5. Results and discussion

In order to assess the efficiency of AR Yule-Walker and AR-FBLP in whitening the EEG background colored noise, 20 EEG data segments, each of 130 samples, are collected and bandpass filtered to 0.03–30 Hz. A sampling time of 3 milliseconds is used. The autocorrelation function of the random process is obtained by using the biased estimator [7]. The FPE criterion is used to estimate the AR model order p in each data segment. After defining the AR(p) coefficients that best model the EEG data segments by using the AR Yule-Walker or AR-FBLP techniques, the data is whitened by passing it through an *error linear predictor* (ELP) filter of similar order and coefficients to AR(p). Time- and frequency-domain analyses are carried out to investigate the whitening potential of the AR Yule-Walker and AR-FBLP parametric techniques.

In time-domain analysis, the EEG colored noise is applied to the input of an ELP filter and the autocorrelation function of the output is calculated and compared with an ideal one (impulse function).

To show the whitening effect of the AR Yule-Walker and AR-FBLP, 20 EEG data sequences are passed through the two ELP filters and the autocorrelation functions of the inputs and outputs are calculated and depicted in Figure 5.1.

Figure 5.1 shows clearly the efficiency of AR Yule-Walker and AR-FBLP algorithms in whitening the EEG colored noise. Both algorithms managed to reduce the arbitrary autocorrelation function of the EEG colored noise, Figure 5.1(a), to approximately the unit impulse of the white noise, Figures 5.1(b) and 5.1(c), with relatively better performance by the AR-FBLP algorithm than by the AR Yule-Walker.

For better view of the performances of the considered whitening algorithms, the frequency domain characteristic of the whitened noise is also obtained and compared.

The spectra of the 20 EEG colored noise sequences are first obtained and depicted in Figure 5.2(a). Next, the spectra of the 20 whitened EEG data sequences at the outputs of
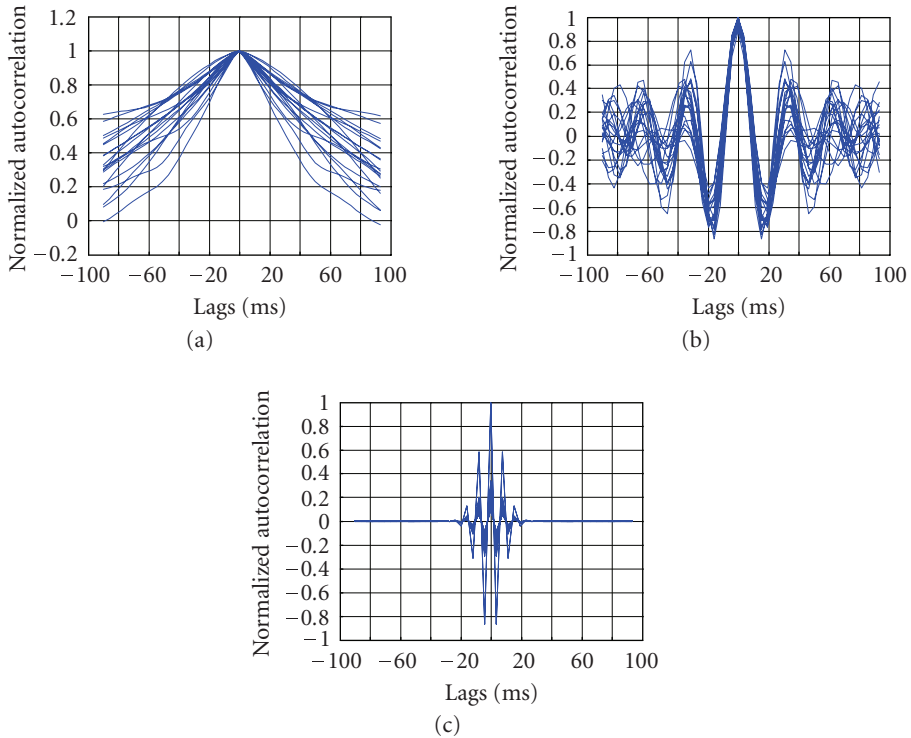
Figure 5.1. (a) The autocorrelation of the EEG colored noise, (b) autocorrelation of the EEG colored noise whitened by AR Yule-Walker, (c) autocorrelation of the EEG colored noise whitened by AR-FBLP.

AR Yule-Walker and AR-FBLP filters are calculated and depicted, respectively, in Figures 5.2(b) and 5.2(c).

The results in Figure 5.2 show clearly the better whitening effect of AR-FBLP than that of AR Yule-Walker, where the frequency characteristics are much closer to the flat spectrum of ideal white noise. It is useful to indicate here that the major source of the ripples in the spectra of the two ELP filters is mainly the limited time extent of EEG data sequences. Accordingly, and in order to flatten more the whitened process, the number of samples of the 20 EEG data sequences is increased from 130 to 256 samples. The spectra are shown in Figures 5.3(a) and 5.3(b).

6. Conclusion

The use of the autoregressive model in whitening the background EEG colored noise in brain activity is discussed. Two different methods for estimating the autoregressive model parameters are proposed and thoroughly investigated. The AR Yule-Walker is proposed as a representative of the second-order moments-based algorithms and the AR forward-backward linear prediction is proposed as a representative of the data-oriented algorithms. 20 EEG data segments are collected when no VEPs brain signals exist and
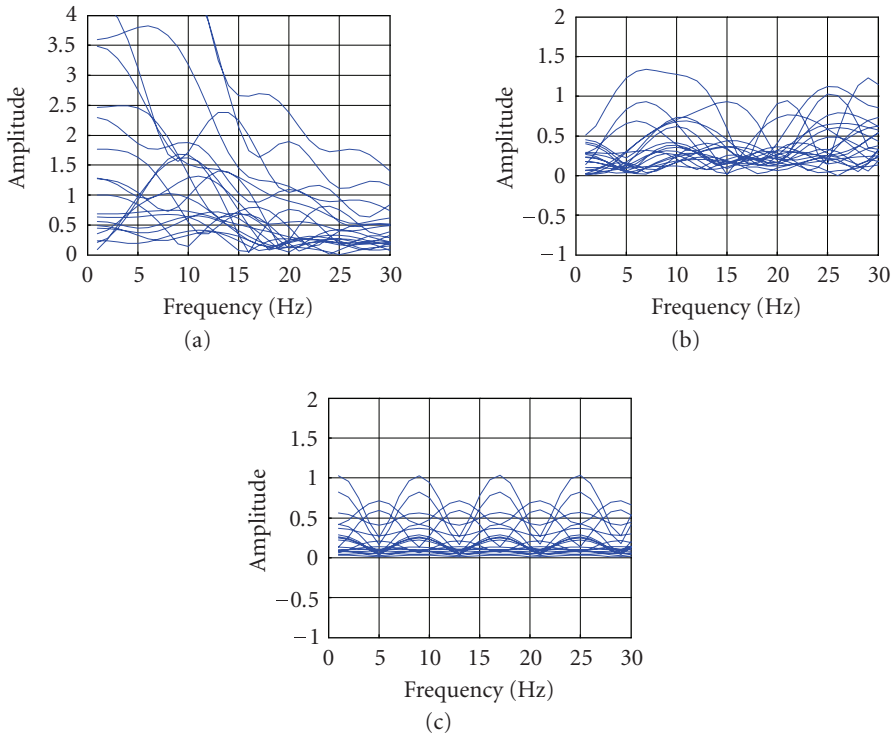
FIGURE 5.2. (a) Spectrum of colored EEG noise, (b) the spectrum of a whitened EEG noise by AR Yule-Walker, (c) the spectrum of a whitened EEG noise by AR-FBLP.
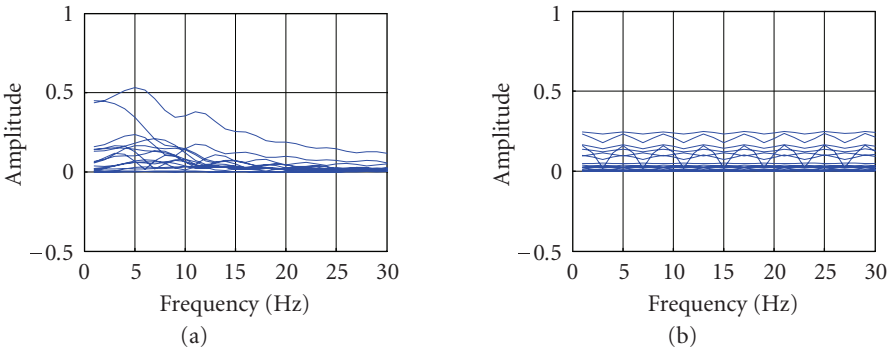

FIGURE 5.3. (a) The spectrum of a whitened EEG noise by AR Yule-Walker, (b) the spectrum of a whitened EEG noise by AR-FBLP.

used to asses the whitening capabilities of the two parametric methods. The performance of the AR Yule-Walker and AR-FBLP in whitening colored noise is explored in time and frequency domains. Both algorithms show good potential in whitening the spectrum of the colored EEG noise with relatively better performance by AR-FBLP.

References

[1] K. H. Chiappa and C. Yiannikas, *Evoked Potentials in Clinical Medicine*, Raven Press, New York, NY, USA, 1983.

[2] Y. Wang and F. Yang, "Dynamic extraction of visual evoked potentials through spatial analysis and dipole localization," *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 8, pp. 762–768, 1995.

[3] J. Makhoul, "Linear prediction: a tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[4] A. Papoulis, "Maximum entropy and spectral estimation: a review," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 6, pp. 1176–1186, 1981.

[5] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1992.

[6] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, NJ, USA, 2003.

[7] S. L. Marple Jr., *Digital Spectral Analysis with Applications*, Prentice Hall Signal Processing Series, Prentice-Hall, Englewood Cliffs, NJ, USA, 1987.

[8] S. Kay, *Modern Spectral Estimation*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1988.

[9] M. Hayes, *Statistical Digital Signal Processing and Modeling*, John Wiley & Sons, New York, NY, USA, 1996.

[10] H. Akaike, "Statistical predictor identification," *Annals of the Institute of Statistical Mathematics*, vol. 22, no. 1, pp. 203–217, 1970.

[11] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

Nidal Kamel: Department of Electrical and Electronic Engineering, Universiti Teknologi Petronas, Bandar Seri Iskandar, Tronoh 31750, Perak, Malaysia
Email address: nidalkamel@petronas.com.my

Andrews Samraj: The Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, Melaka 75450, Malaysia
Email address: andrews.samraj@mmu.edu.my

Arash Mousavi: The Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, Melaka 75450, Malaysia
Email address: arash.mousavi@mmu.edu.my